

# Identification of Stochastic Wiener Systems using Indirect Inference<sup>\*</sup>

Bo Wahlberg<sup>\*</sup> James Welsh<sup>\*\*</sup> Lennart Ljung<sup>\*\*\*</sup>

<sup>\*</sup> *Department of Automatic Control and ACCESS, School of Electrical Engineering, KTH Royal Institute of Technology, SE-100 44 Stockholm, Sweden. (e-mail: bo.wahlberg@ee.kth.se).*

<sup>\*\*</sup> *School of Elect Engineering and Computer Science, The University of Newcastle, Callaghan NSW 2308, Australia.*

<sup>\*\*\*</sup> *Division of Automatic Control, Linköping University, SE-581 83 Linköping, Sweden.*

**Abstract:** We study identification of stochastic Wiener dynamic systems using so-called indirect inference. The main idea is to first fit an auxiliary model to the observed data and then in a second step, often by simulation, fit a more structured model to the estimated auxiliary model. This two-step procedure can be used when the direct maximum-likelihood estimate is difficult or intractable to compute. One such example is the identification of stochastic Wiener systems, i.e., linear dynamic systems with process noise where the output is measured using a non-linear sensor with additive measurement noise. It is in principle possible to evaluate the log-likelihood cost function using numerical integration, but the corresponding optimization problem can be quite intricate. This motivates studying consistent, but sub-optimal, identification methods for stochastic Wiener systems. We will consider indirect inference using the best linear approximation as an auxiliary model. We show that the key to obtain a reliable estimate is to use uncertainty weighting when fitting the stochastic Wiener model to the auxiliary model estimate. The main technical contribution of this paper is the corresponding asymptotic variance analysis. A numerical evaluation is presented based on a first-order finite impulse response system with a cubic non-linearity, for which certain illustrative analytic properties are derived.

**Keywords:** system identification, indirect inference, identification of non-linear systems, Wiener systems, stochastic non-linear system.

## 1. INTRODUCTION

The idea of using two-stage identification methods, e.g., indirect inference, is by no means new in system identification. Typically, a flexible auxiliary model is fitted to data, and then in a second step this estimated model is used to find a more structured model. A well-known example of such an approach is the indirect Prediction Error Minimization (PEM) method, Söderström et al. (1991), where it is assumed that the model structure of interest can be embedded in a larger model structure for which the identification problem is more tractable. In a second step, the structured model is estimated from the larger model using a weighted non-linear least squares method. The indirect PEM method will, under certain assumptions, have the same asymptotic statistical properties as the Maximum-Likelihood (ML) method, but it can be more efficiently calculated. Indirect PEM is a special case of indirect inference, which was introduced in econometrics in Gouriéroux et al. (1993). Their main motivation was identification problems for which the ML method is in-

tractable. They also proposed the use of Monte Carlo simulations to generate the cost to be minimized in the second step.

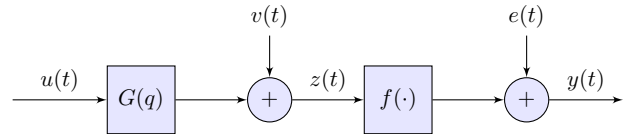


Fig. 1. Stochastic Wiener System

The concept of indirect inference was introduced to the system identification community by Welsh et al. (2009) and Larsson et al. (2010). The aim of the current paper is to provide further insights into the potential use of indirect inference for the identification of scalar discrete time stochastic Wiener systems, illustrated in Figure 1, of the form

$$\begin{aligned} z(t) &= G(q)u(t) + v(t), \\ y(t) &= f(z(t)) + e(t), \end{aligned} \quad (1)$$

with a stable transfer function  $G(q)$  (where  $q$  denotes the shift operator), an input signal sequence  $\{u(t)\}$ , white stationary process noise  $\{v(t)\}$  with zero mean and variance  $\sigma_v^2$ , an output signal sequence  $\{y(t)\}$ , and additive white stationary measurement noise  $\{e(t)\}$  with zero mean and variance  $\sigma_e^2$ . The input signal  $u(t)$  is assumed to

<sup>\*</sup> This work was partially supported by the Swedish Research Council and the Linnaeus Center ACCESS at KTH. The research leading to these results has received funding from The European Research Council under the European Community's Seventh Framework program (FP7 2007-2013) / ERC Grant Agreement N. 267381

be independent of the noises (the open-loop case), and may be a realization of a stochastic process with known probability distribution. To simplify the presentation, we will assume that the noise processes are independent and normal (gaussian) distributed. The more general case with coloured process noise can be handled using a predictor model.

The main challenge is the non-linear function  $f(\cdot)$ , which means that we have a non-linear stochastic system where the process noise  $\{v(t)\}$  propagates through a non-linear device. This typically corresponds to a non-linear sensor.

### 1.1 Contribution

The main objective of this paper is to study the indirect inference method using the best linear approximation as an auxiliary model and compare this to maximum-likelihood and prediction error minimization methods for identification of a stochastic Wiener system. The main challenge is to handle the non-linear process noise contribution. It is in principle possible to calculate the likelihood cost, or some approximation of it, using for example particle filters and/or Markov Chain Monte Carlo methods, but even for simple examples the maximum-likelihood approach leads to rather involved computations. A problem for prediction error minimization methods is that the probability density function of the conditional mean prediction error is quite complicated and involves convolution integrals. This is the reason for using more ad hoc identification techniques. A common approach is to use linear and gaussian approximations. The main technical contributions are:

- To connect the use of best linear approximation of stochastic Wiener systems with the method of indirect inference. We present the corresponding variance uncertainty weighting, which for this case includes the input signal. We also derive the asymptotic variance expression for the estimated model parameters.
- Illustration of the results on a first-order finite impulse response model, for which it is possible to find analytic expressions. The statistical performance on this example of indirect inference based on best linear approximation is comparable to that of maximum-likelihood estimation and prediction error minimization. The computations for the indirect inference method are just a fraction of the ones for calculating the maximum-likelihood estimate. The example also shows that the cost of using a non-linear sensor is increased uncertainty in the estimated model.

The statistical theory for identification of stochastic Wiener systems is by no means complete, and the aim of this paper is to provide insights in some open important problems.

### 1.2 System Identification

Given measurements of the input and output signals  $\{u(t), y(t)\}$ ,  $t = 1, \dots, N$ , the task is to identify a model of the stochastic Wiener system of the form,

$$\begin{aligned} z(t) &= G(q, \theta)u(t) + v(t), \\ y(t) &= f(z(t)) + e(t) \end{aligned} \quad (2)$$

where the model is parameterized by the parameter vector  $\theta \in \mathbf{R}^n$ . We assume that the true system can be described by  $\theta_o$ . The noise processes  $\{v(t)\}$  and  $\{e(t)\}$  are assumed to be independent normal distributed (gaussian) zero mean white noise. The corresponding noise variances  $\sigma_v^2$  and  $\sigma_e^2$  are assumed to be known, but could be added to the parameters to be estimated. This can, however, cause identifiability problems. We will study the case when the non-linear function  $f(\cdot)$  is known. It is possible to extend our results to the case when the function  $f(\cdot)$  also is estimated, which can result in an identifiability problem. The reason for these simplifications is to focus on the stochastic part due to the process noise  $v(t)$ .

Identification of Wiener systems is a well studied topic, see for example Greblicki (1992); Wigren (1993); Bai (2003); Zhu (2002); Enqvist and Ljung (2005); Pillonetto (2013); Giri and Bai (2010) and the references therein. It forms the basis for the identification of more general non-linear block diagram based models. However, many algorithms assume no process noise, which leads to a non-linear least squares problem minimizing the output error, i.e., the difference between the measured and the simulated outputs. The maximum-likelihood method for stochastic Wiener systems was introduced in Hagenblad and Ljung (2000) and analysed in more detail in Hagenblad et al. (2008). The expectation-maximization algorithm for maximum-likelihood identification and the use of the particle filter have been studied in Wills et al. (2013); Wills and Ljung (2010). As recently pointed out in Wahlberg et al. (2014), the stochastic Wiener system identification problem can be viewed as a non-linear errors-in-variables problem, with the well-known bias problem due to input noise. This paper also proposes a prediction error minimization framework for identification of stochastic Wiener systems. The idea is to use the conditional mean predictor and notice that the variance of the prediction errors may be highly dependent on the input signal  $u(t)$ . Hence, variance uncertainty weighting is most important in order to obtain reliable estimates.

### 1.3 Outline

ML and PEM methods for stochastic Wiener system identification are reviewed in Sec. 2. In Sec. 3, we study how to use the indirect inference approach for the identification of stochastic Wiener models by using the BLA as the auxiliary model. A first order FIR model example is outlined in Sec. 4, and the corresponding numerical study is presented in Sec. 5. The paper is concluded in Sec. 6.

## 2. ML AND PEM

As shown in Hagenblad and Ljung (2000), the negative log-likelihood function, given data and the normal distributed noise model (2), equals

$$l(\theta) = - \sum_{t=1}^N \log \int_{-\infty}^{\infty} e^{-\mathcal{E}(t, \theta, z)} dz,$$

where

$$\mathcal{E}(t, \theta, z) = \frac{[y(t) - f(z)]^2}{2\sigma_e^2} + \frac{[z - G(q, \theta)u(t)]^2}{2\sigma_v^2}.$$

The ML estimate of  $\theta$  is obtained by minimizing  $l(\theta)$ . There are at least two challenges with the ML method for

stochastic Wiener systems. First, to evaluate the negative log-likelihood cost at a certain parameter value  $\theta$  we have to calculate  $N$  integrals. This can be done rather efficiently using numerical integration and parallel computations. A difficulty is the integrand, where typically

$$e^{-\mathcal{E}(t, \theta, x)} \approx \begin{cases} 1, & x \text{ small,} \\ 0, & x \text{ otherwise.} \end{cases}$$

The example to be considered in Sec. 4 corresponds to  $\mathcal{E}(x) \sim x^6$ , which means that the integrand decreases very rapidly to zero.

The PEM approach avoids the exponential function integration issue by using a weighted least squares cost function, see Wahlberg et al. (2014). The conditional mean predictor of  $y(t)$  for given  $u(t)$  and  $\theta$  is

$$\hat{y}(t, \theta) = E_v\{f(G(q, \theta)u(t) + v(t))\}. \quad (3)$$

Notice that the prediction error variance depends on the input signal  $u(t)$ . The optimally weighted quadratic PEM cost-function, see Wahlberg et al. (2014), is

$$V_N(\theta) = \frac{1}{N} \sum_{t=1}^N \frac{\epsilon^2(t, \theta)}{E\{\epsilon^2(t, \hat{\theta}_I)\}}, \quad (4)$$

with prediction error  $\epsilon(t, \theta) = y(t) - \hat{y}(t, \theta)$ . The variance weighting is calculated at a consistent initial estimate  $\hat{\theta}_I$ , e.g., the PEM estimate without weighting. The use of weighting is important to obtain reliable estimates and depends here on the input signal  $u(t)$ . The PEM estimate based on (4) is not asymptotically efficient for non-linear functions, since we use a weighted quadratic cost-function. However, by using a cost-function based on the probability density function of  $\epsilon(t, \theta)$ , we obtain an asymptotically efficient PEM estimate, see Ljung (1999). The computations will then be similar to the ML case, involving multiple integral calculations with exponential functions.

### 3. INDIRECT INFERENCE USING BEST LINEAR APPROXIMATION

#### 3.1 Best Linear Approximation of Stochastic Wiener Systems

Let us illustrate the concept of indirect inference applied to a stochastic Wiener system with a known non-linear function. A common first approach is to fit a linear model to the observed data. It is well known that if the input signal is normal (gaussian) distributed, then the Best Linear Approximation (BLA), see Ljung (2001) or (Giri and Bai, 2010, Chapter 13), is a scaled version of the linear dynamics transfer function  $G(q)$  of the Wiener system. It is perhaps less well known that the same result holds for *stochastic* Wiener systems with gaussian process noise  $v(t)$ . This extension follows more or less from Bussgang's theorem, Bussgang (1952): If  $z(t)$  is a normal distributed stationary stochastic process with zero mean and if the non-linear transformed process  $y(t) = f(z(t))$  has zero mean, then

$$E\{y(t)z(t - \tau)\} = b_0 E\{z(t)z(t - \tau)\}, \quad b_0 = E\{f'(z(t))\}.$$

Adding independent gaussian process noise to the input signal contribution,  $z(t) = G(q)u(t) + v(t)$ , makes  $z(t)$  still normal distributed and Bussgang's theorem holds. Now we

are only interested in computing the partial correlations using the relations

$$\begin{aligned} E\{y(t)u(t - \tau)\} &= b_0 E\{z(t)u(t - \tau)\} \\ &= b_0 E\{[G(q)u(t)]u(t - \tau)\}. \end{aligned} \quad (5)$$

A recent proof of (5) can be found in Banelli (2012), but it already follows from Nuttall (1958). This result proves that the BLA of a stochastic Wiener system with a normal distributed input signal again equals

$$G_{BLA}(q) = b_0 G(q)$$

i.e., a scaled version of the linear transfer function. Notice that  $b_0$  now also depends on the statistics of the process noise  $v$  as well as of the input  $u$ . This result can be generalized to separable stochastic processes, Nuttall (1958); Enqvist and Ljung (2005).

The BLA of a stochastic Wiener system for more general input signal sequences can be obtained by simulations or by analytic calculations (if the distribution of the input signal is known).

#### 3.2 An Optimally Weighted Indirect Inference Algorithm

We will now describe how to apply indirect inference to the stochastic Wiener system identification problem using a two-step procedure based on BLA. From the given data  $\{u(t), y(t)\}$ ,  $t = 1, \dots, N$ , form the BLA cost-function

$$Q_N(\beta) = \frac{1}{N} \sum_{t=1}^N [y(t) - G_{lin}(q, \beta)u(t)]^2. \quad (6)$$

Here we have used the notation  $G_{lin}(q, \beta)$  to stress that this is a linear transfer function parameterized by  $\beta \in \mathbf{R}^m$ . It is in general of higher order compared to  $G(q, \theta)$  in (2), that is  $m$  is typically larger than  $n$ , the dimension of  $\theta$ . We have used an Output Error PEM cost-function (6), but it is also possible to use a more general PEM model structure, as described in Ljung (2001).

**Step 1:** Identify the BLA within the model structure  $G_{lin}(q, \beta)$ ,  $\beta \in \mathbf{R}^m$ , of the system from given data using the cost-function (6)

$$\hat{\beta}_N = \arg \min_{\beta} \{Q_N(\beta)\},$$

i.e., the BLA estimate equals  $\hat{G}_{BLA}(q) = G_{lin}(q, \hat{\beta}_N)$ .

The next question is to figure out the functional relation  $\beta(\theta)$ ,  $\mathbf{R}^m \mapsto \mathbf{R}^n$  that describes how the auxiliary estimate  $\hat{\beta}_N$  asymptotically depends on the underlying structured model of interest, as a function of the model parameter vector  $\theta$ . Now assume that  $Q_N(\beta)$  converges (almost surely) as the number of data  $N \rightarrow \infty$  to  $Q(\beta, \theta_o)$ , which depends on the true system parameter vector  $\theta_o$ . Here

$$\begin{aligned} Q(\beta, \theta) &= E\{[y(t) - G_{lin}(q, \beta)u(t)]^2\} \\ &= E_{v,u}\{[f(G(q, \theta)u(t) + v(t)) - G_{lin}(q, \beta)u(t)]^2\}. \end{aligned}$$

The expectation is with respect to both the input signal and the process noise. It is also possible to only use expectation with respect to the process noise for a given input sequence. Let

$$\beta(\theta) = \arg \min_{\beta} \{Q(\beta, \theta)\},$$

and use the notation

$$\beta_o = \beta(\theta_o),$$

for the corresponding true parameter vector. This leads to:

**Step 2: (Analytic)** Estimate the structured model parameter vector  $\theta$  by solving

$$\hat{\theta}_N = \arg \min_{\theta} [\beta(\theta) - \hat{\beta}_N]^T W [\beta(\theta) - \hat{\beta}_N], \quad (7)$$

where  $W$  is a positive definite weighting matrix to be specified.

For certain examples of non-linear functions and distributions it may not be possible to analytically find the function  $\beta(\theta)$ . One can then resort to Monte Carlo simulations. Let

$$Q_{N,S}(\beta, \theta) = \frac{1}{S} \sum_{s=1}^S \frac{1}{N} \sum_{t=1}^N [f(G(q, \theta)u(t) + v_s(t)) - G_{lin}(q, \beta)u(t)]^2.$$

Here  $\{v_s(t)\}$ ,  $t = 1, \dots, N$ ,  $s = 1, \dots, S$ , is a generated noise realization of the process noise  $v(t)$  and  $S$  is the total number of realizations used in the Monte Carlo Simulation. Let

$$\hat{\beta}_{N,S}(\theta) = \arg \min_{\beta} \{Q_{N,S}(\beta, \theta)\}.$$

**Step 2: (Simulated)** Estimate the structured model parameter vector  $\theta$  by solving

$$\hat{\theta}_{N,S} = \arg \min_{\theta} [\hat{\beta}_{N,S}(\theta) - \hat{\beta}_N]^T W [\hat{\beta}_{N,S}(\theta) - \hat{\beta}_N], \quad (8)$$

where  $W$  is a positive definite weighting matrix to be specified.

The *optimal weighting matrix*,  $W$ , equals the inverse of the covariance matrix of the auxiliary estimate  $\hat{\beta}_N$ , compare Expression (9.11) for the asymptotic covariance matrix in Ljung (1999). In our setting this translates to

$$W = [J_o^{-1} I_o J_o^{-1}]^{-1}, \quad (9)$$

$$I_o = \lim_{N \rightarrow \infty} \text{Cov}\{\sqrt{N} \frac{\partial Q_N}{\partial \beta}(\beta_o, \theta_o)\}, \quad J_o = \frac{\partial^2 Q}{\partial \beta^2}(\beta_o, \theta_o).$$

In practise, one has to use a consistent estimate of  $W$ .

### 3.3 Performance Analysis

Next, we determine the asymptotic properties of the final estimate  $\hat{\theta}_N$ . This can also be done using Taylor series expansion arguments as described in Complement C4.4 in Söderström and Stoica (1989). The properties of the function  $\beta(\theta)$ ,  $\mathbf{R}^m \mapsto \mathbf{R}^n$ , are very important in order to obtain a consistent estimate. It should be possible to invert this function,  $\theta = \alpha(\beta(\theta))$ , i.e.,  $\alpha(\cdot)$  is a left inverse of  $\beta(\cdot)$ . Define the Jacobian matrix

$$G = \frac{\partial \beta}{\partial \theta}(\theta_o) \in \mathbf{R}^{m \times n}, \quad (10)$$

and recall that we use the weighting matrix

$$W = [\lim_{N \rightarrow \infty} \text{Cov}\{\sqrt{N}[\hat{\beta}_N - \beta_o]\}]^{-1}. \quad (11)$$

*Variance Expression:* The asymptotic covariance matrix of  $\hat{\theta}_N$  defined by (7) equals

$$\lim_{N \rightarrow \infty} \text{Cov}\{\sqrt{N}[\hat{\theta}_N - \theta_o]\} = [G^T W G]^{-1}, \quad (12)$$

where  $G$  is defined by (10) and  $W$  by (11).

To prove this result, we will make a Taylor series expansion of the cost-function at the minimizing  $\theta$ :

$$V_w(\theta) = [\beta(\theta) - \hat{\beta}_N]^T W [\beta(\theta) - \hat{\beta}_N] \Rightarrow V'_w(\hat{\theta}_N) = 0,$$

which gives  $[\hat{\theta}_N - \theta_o] \approx -[V''_w(\theta_o)]^{-1} V'_w(\theta_o)$ . The Hessian  $V''_w(\theta_o)$  has to be invertible (positive definite for unique

local minimum) for this to hold. As shown in Complement 4.4 in Söderström and Stoica (1989)

$$V'_w(\theta_o) = 2G^T W [\hat{\beta}_N - \beta_o] + \mathcal{O}(1/N)$$

$$V''_w(\theta_o) = 2G^T W G + \mathcal{O}(1/\sqrt{N}).$$

Hence

$$[\hat{\theta}_N - \theta_o] \approx -[G^T W G]^{-1} G^T W [\hat{\beta}_N - \beta_o].$$

Since  $\lim_{N \rightarrow \infty} \text{Cov}\{\sqrt{N}[\hat{\beta}_N - \beta_o]\} = W^{-1}$ , this proves (12).

When using the Monte Carlo simulation approach in Step 2, (8), the asymptotic covariance matrix of  $\hat{\theta}_{N,S}$  should be amplified by the factor  $(1 + 1/S)$ , due to the extra uncertainty from the simulations, see Heggland and Frigessi (2004).

### 3.4 Comments

The suggestion to use the BLA as an auxiliary model is rather ad hoc, but a very common choice in recent methods for identification of non-linear systems, Schoukens et al. (2005); Pintelon and Schoukens (2012); Schoukens et al. (2014); Schoukens and Rolain (2012); Sjöberg and Schoukens (2012). Ljung (2001) contains an overview of the role of BLA in system identification. Another option would be to use the biased estimate from minimizing

$$l(\beta) = \sum_{t=1}^N [y(t) - f(G(q, \beta)u(t))]^2$$

and then use Step 2 to remove the bias of this estimate. A challenge is to find even more efficient auxiliary models. The key property is that  $\beta(\theta)$  should have a left inverse (identifiability) and

$$G = \frac{\partial \beta}{\partial \theta}(\theta_o) \in \mathbf{R}^{m \times n} \quad (13)$$

should be “large” (sensitive) and at the same time  $\beta$  should be easy to estimate from the given data.

### 3.5 Indirect Inference

As mentioned in the introduction the indirect inference approach was developed in Gourieroux et al. (1993) as a way to find consistent model estimates even when the ML method is intractable. Our description of the indirect inference approach applied to BLA in the previous section is mainly based on Heggland and Frigessi (2004). The theory of indirect inference is more general, and the key step is to choose the auxiliary model parameterized by  $\beta$  and the data-driven cost-function  $Q_N(\beta)$ . A convergence analysis of the general indirect inference method presented in Heggland and Frigessi (2004) is based on using the concept of finite dimension auxiliary statistics.

If  $Q_N(\beta)$  is based on a *sufficient statistics* the indirect inference method is efficient (except for the factor  $(1 + 1/S)$ ). This is not the case for stochastic Wiener systems, for which no finite dimensional sufficient statistics exists. For the BLA approach we are only using a second order statistics.

## 4. ILLUSTRATING EXAMPLE

We will use the following simple example to try to further understand the properties of the stochastic Wiener system identification methods described in the previous sections,

$$\begin{aligned} z(t) &= \theta u(t) + u(t-1) + v(t) \\ y(t) &= [z(t)]^3 + e(t). \end{aligned} \quad (14)$$

The same example was used in Larsson et al. (2010); Wahlberg et al. (2014). We will study two different input white noise distributions, a normal (gaussian) and an uniform distribution, both with variance  $\sigma_u^2$ . The noises are assumed to be white zero mean normal distributed with variances  $\sigma_e^2$  and  $\sigma_v^2$ , respectively.

The maximum-likelihood cost-function

$$l(\theta) = - \sum_{t=1}^N \log \mathbb{E}_{\bar{v}} \left\{ \frac{\sigma_v}{\sqrt{2\pi}} e^{-\frac{1}{2\sigma_e^2} [y(t) - (\theta u(t) + u(t-1) + \sigma_v \bar{v})]^3} \right\}$$

is calculated using a Gauss Hermite approximation of order 1000. The reason for this high order is the  $e^{-x^6}$  tends to zero very quickly and the integral is rather difficult to approximate. We have noticed that increased process noise variance  $\sigma_v^2$  makes it even more challenging.

The conditional mean predictor equals

$$\begin{aligned} \hat{y}(t, \theta) &= \theta^3 u^3(t) + u^3(t-1) + 3(\theta u^2(t)u(t-1) \\ &\quad + \theta u(t)u^2(t-1) + \theta u(t)\sigma_v^2 + u(t-1)\sigma_v^2), \end{aligned} \quad (15)$$

and is used in the PEM method (4).

In order to evaluate the indirect inference method we will study the BLA. Assume first that the input signal is normal distributed with zero mean and variance  $\sigma_u^2$ . The BLA of model order one, i.e., based on minimizing

$$\mathbb{E}\{(y(t) - \beta_1 u(t) + \beta_2 u(t-1))^2\}$$

is

$$\begin{pmatrix} \beta_1(\theta) \\ \beta_2(\theta) \end{pmatrix} = [3\sigma_u^2\theta^2 + 3(\sigma_u^2 + \sigma_v^2)] \begin{pmatrix} \theta \\ 1 \end{pmatrix}, \quad (16)$$

for which

$$\frac{\beta_1}{\beta_2} = \theta,$$

since the BLA for gaussian input signal is just a scaled version of  $G(q, \theta)$ . This also gives that  $\beta(\alpha)$  has a left inverse and we have identifiability of  $\theta$  from  $\beta$ .

The corresponding BLA when the input signal is uniformly distributed with zero mean variance  $\sigma_u^2$  is

$$\begin{pmatrix} \beta_1(\theta) \\ \beta_2(\theta) \end{pmatrix} = \begin{pmatrix} \frac{9}{5}\sigma_u^2\theta^3 + 3(\sigma_u^2 + \sigma_v^2)\theta \\ 3\sigma_u^2\theta^2 + 3(\frac{3}{5}\sigma_u^2 + \sigma_v^2) \end{pmatrix}. \quad (17)$$

For the indirect inference method we will use

$$Q_N(\beta) = \frac{1}{N} \sum_{t=1}^N (y(t) - \beta_1 u(t) + \beta_2 u(t-1))^2$$

and hence

$$\frac{\partial Q_N}{\partial \beta} = \frac{2}{N} \sum_{t=1}^N \begin{pmatrix} (y(t) - \beta_1 u(t) + \beta_2 u(t-1))u(t) \\ (y(t) - \beta_1 u(t) + \beta_2 u(t-1))u(t-1) \end{pmatrix}$$

and we need to calculate the weighting matrix  $W$  from (9). Here  $J_o = \sigma_u^2 I$  and the tedious work is to calculate  $I_o$ . It can, however, be estimated as

$$\begin{aligned} \hat{I}_o &= \frac{1}{N} \sum_{t=1}^N \left[ [y(t) - \hat{\beta}_1 u(t) - \hat{\beta}_2 u(t-1)]^2 \times \right. \\ &\quad \left. \begin{pmatrix} u^2(t) & u(t)u(t-1) \\ u(t)u(t-1) & u^2(t-1) \end{pmatrix} \right] \end{aligned}$$

**Comment:** The key to get the order one indirect inference method to work is to use the weighting  $W = \hat{I}_o^{-1}$ . The paper Larsson et al. (2010) does not use weighting, which explains their non-intuitive simulation result that it is better to use a zero order BLA model than a first order one (which should contain more information).

We will finally study the zero order model case, for which no weighting is needed. Here

$$Q_N(\beta) = \frac{1}{N} \sum_{t=1}^N (y(t) - \beta_1 u(t))^2$$

with the BLA given by  $\beta_1(\theta)$  in (16) and (17). An alternative cost-function is

$$Q_N(\beta) = \frac{1}{N} \sum_{t=1}^N (y(t) - \beta_2 u(t-2))^2,$$

with the BLA given by  $\beta_2(\theta)$  from (16) or (17). A problem here is that  $\beta_2(\theta)$  is quadratic in  $\theta$  and we do not have an unique solution with respect to  $\theta$ .

## 5. SIMULATION RESULT

We will use the following numerical values for the system parameters in (14):

$$\theta_o = 0.5, \quad \sigma_v^2 = 0.2, \quad \sigma_e^2 = 0.1.$$

We will use the analytic Step 2, (7), for the indirect inference based on BLA, and will evaluate the following methods for  $N = 1000$  observations:

- Method 1: ML
- Method 2: PEM with optimal weighting
- Method 3: Zero order indirect inference
- Method 4: First order indirect inference, no weighting
- Method 5: First order indirect inference, with weighting

The following table summarizes the simulation results for the *normal distributed input sequence* with zero mean and variance  $\sigma_u^2 = 1/3$  over 1000 noise and input realizations:

Method 1:	mean: 0.5025	std: 0.0303
Method 2:	mean: 0.5014	std: 0.0349
Method 3:	mean: 0.4983	std: 0.0446
Method 4:	mean: 0.4977	std: 0.0554
Method 5:	mean: 0.4982	std: 0.0418

The next table summarizes the simulation results for the *uniform distributed input sequence* with zero mean and variance  $\sigma_u^2 = 1/3$  over 1000 noise and input realizations

Method 1:	mean: 0.4994	std: 0.0325
Method 2:	mean: 0.4984	std: 0.0346
Method 3:	mean: 0.4988	std: 0.0454
Method 4:	mean: 0.4984	std: 0.0458
Method 5:	mean: 0.4987	std: 0.0377

The conclusion from the numerical study is that the ML, as expected, gives the best performance. However, PEM and indirect inference with optimal weighting give also good

results. The computational times for these methods are only a fraction of that of ML. The simulation study also shows that a zero order model can give better results than using a first order model and no weighting, c.f., Larsson et al. (2010).

It should be noted that the asymptotic variance when  $f(x) = x$  equals  $(\sigma_v^2 + \sigma_e^2)/(\sigma_u^2 N)$ , with a standard deviation equal to 0.0173 for the example above. Hence, the non-linear function  $f(x) = x^3$  gives a more difficult estimation problem than the standard case  $f(x) = x$ .

## 6. CONCLUSION

In this paper, we have utilised the indirect inference method based on the best linear approximation to identify stochastic Wiener systems. The results show that to obtain a reliable estimate, it is important to use an optimal weighting when estimating the structured model from the auxiliary model. The weighting here is the inverse of the covariance matrix of the BLA parameter estimate. We have analyzed the statistical properties of the corresponding indirect inference estimate using results from system identification. The methods have been evaluated using a first order FIR system with a cubic non-linearity. The simulation results demonstrate that the proposed indirect inference BLA approach performs quite well compared to the ML and weighted PEM methods. A major advantage is that the indirect inference algorithms are very computationally fast and direct to implement.

There are many open questions when it comes to identification of stochastic non-linear dynamic systems. For example, performance results to guide the choice of sensors.

## REFERENCES

- Bai, E.W. (2003). Frequency domain identification of Wiener models. *Automatica*, 39(9), 1521–1530.
- Banelli, P. (2012). Non-linear transformations of gaussians and gaussian-mixtures with implications on estimation and information theory. *IEEE Trans. on Inf. Theory*. submitted, available on arXiv:1111.5950v3 [cs.IT].
- Bussgang, J. (1952). Cross-correlation function of amplitude-distorted gaussian signals. Technical Report 216, MIT Laboratory of Electronics.
- Enqvist, M. and Ljung, L. (2005). Linear approximations of nonlinear FIR systems for separable input processes. *Automatica*, 41(3), 459–473.
- Giri, F. and E-W. Bai, (Eds.) (2010). *Block-oriented Non-linear System Identification*. Lecture Notes in Control and Information Sciences. Volume 404, ISBN: 978-1-84996-512-5.
- Gourieroux, C., Monfort, A., and Renault, E. (1993). Indirect inference. *Journal of applied econometrics*, 8(S1), S85–S118.
- Greblicki, W. (1992). Nonparametric identification of Wiener systems. *IEEE Transactions on Information Theory*, 38(5), 1487–1493.
- Hagenblad, A. and Ljung, L. (2000). Maximum likelihood estimation of Wiener models. In *Proceedings of the 39th IEEE Conference on Decision and Control, 2000*, volume 3, 2417–2418 vol.3.
- Hagenblad, A., Ljung, L., and Wills, A. (2008). Maximum likelihood identification of Wiener models. *Automatica*, 44(11), 2697–2705.
- Heggland, K. and Frigessi, A. (2004). Estimating functions in indirect inference. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 66(2), 447–462.
- Larsson, C., Hjalmarsson, H., and Rojas, C. (2010). Identification of nonlinear systems using misspecified predictors. In *49th IEEE Conference on Decision and Control (CDC), 2010*, 7214–7219.
- Ljung, L. (2001). Estimating linear time invariant models of non-linear time-varying systems. *European Journal of Control*, 7(2-3), 203–219. Semi-plenary presentation at the European Control Conference, Sept 2001.
- Ljung, L. (1999). *System Identification: Theory for the User*. Pearson Education.
- Nutall, A. (1958). Theory and application of the separable class of random processes. Technical report, MIT.
- Pillonetto, G. (2013). Consistent identification of Wiener systems: A machine learning viewpoint. *Automatica*, 49(9), 2704 – 2712.
- Pintelon, R. and Schoukens, J. (2012). *System identification: a frequency domain approach*. John Wiley & Sons.
- Schoukens, J., Pintelon, R., Dobrowiecki, T., and Rolain, Y. (2005). Identification of linear systems with nonlinear distortions. *Automatica*, 41(3), 491–504.
- Schoukens, M., Pintelon, R., and Rolain, Y. (2014). Identification of Wiener - Hammerstein systems by a non-parametric separation of the best linear approximation. *Automatica*, 50(2), 628–634.
- Schoukens, M. and Rolain, Y. (2012). Parametric identification of parallel Wiener systems. *IEEE Transactions on Instrumentation and Measurement*, 61(10), 2825–2832.
- Sjöberg, J. and Schoukens, J. (2012). Initializing Wiener Hammerstein models based on partitioning of the best linear approximation. *Automatica*, 48(2), 353 – 359.
- Söderström, T. and Stoica, P. (1989). *System Identification*. Prentice-Hall International.
- Söderström, T., Stoica, P., and Friedlander, B. (1991). An indirect prediction error method for system identification. *Automatica*, 27(1), 183–188.
- Wahlberg, B., Welsh, J.S., and Ljung, L. (2014). Identification of Wiener systems with process noise is a nonlinear errors-in-variables problem. In *53th IEEE Conference on Decision and Control (CDC)*.
- Welsh, J., Aguero, J.C., and Alamir, M. (2009). Continuous-time system identification using indirect inference. In *SYSID 2009*, volume 15, 1169–1174.
- Wigren, T. (1993). Recursive prediction error identification using the nonlinear Wiener model. *Automatica*, 29(4), 1011–1025.
- Wills, A. and Ljung, L. (2010). Wiener system identification using the maximum likelihood method. In *Block-oriented nonlinear system identification*, 89–110. Springer.
- Wills, A., Schön, T.B., Ljung, L., and Ninness, B. (2013). Identification of Hammerstein - Wiener models. *Automatica*, 49(1), 70–81.
- Zhu, Y. (2002). Estimation of an N-L-N Hammerstein - Wiener model. *Automatica*, 38(9), 1607–1614.